

画像観察時のfMRI脳活動パターンを用いた物体姿勢パラメータの推定

鈴木勇翔¹, 田中美里^{2,3}, 青木俊太郎^{2,3}, 澤山 正貴¹, 神谷之康^{2,3}, 土橋宜典^{1,4}

Yuto SUZUKI¹, Misato TANAKA^{2,3}, Shuntaro C. Aoki^{2,3}, Masataka Sawayama¹, Yukiyasu KAMITANI^{2,3}, Yoshinori DOBASHI^{1,4}

¹北海道大学, ²ATR, ³京都大学, ⁴プロメテック CG リサーチ

¹ Hokkaido University, ²ATR, ³Kyoto University, ⁴Prometech CG Research

1 はじめに

近年, fMRI で測定した脳活動をパターン解析することで, 被験者が観察した画像を再構成する試みが行われている. 高精細な画像復元を目指す研究が多くなされる一方で, 画像の根底にある実世界の情報を, 脳活動からどの程度推定できているかについては不明な点が多い. 本研究では, 順光学過程によって画像生成を行う際の物理パラメータを脳活動によってどの程度推定できるかを探ることで, この問題を扱う. 具体的には, CG 画像を観察している際の脳活動を fMRI により測定したデータを用い, 脳活動から再構成した画像と最も近い画像特徴を持つ物体の姿勢を推定した.

2 先行研究

Horikawa らは fMRI により計測された脳活動パターンを深層ニューラルネットワーク (DNN) の特徴量に変換する手法 [1] を提案し, 被験者が観察または想像した画像のカテゴリ予測を実現した. この変換技術を用いて, 脳活動から画像を再構成したのが Shen らによる手法 [2] である. 同手法では, 脳活動から得られる DNN 特徴量と, 入力画像から抽出した DNN 特徴量との誤差が小さくなるように最適化を行い, 画像の再構成を可能にしている. 本研究では, これらの技術をコンピュータグラフィックスに応用し, 脳活動から物体の回転角度に関するパラメータ値を推定する.

3 実験手法

実験手法の概要を図 1 に示す. 図の上段は, 脳活動データから特徴量を得るまでの流れを表している. 本研究では, Horikawa ら [1] および Shen ら [2] の研究と同じ手法を用い, 自然画像を観察中の被験者 1 名から fMRI データを計測し (ボクセルサイズ 2 mm, TR 2 秒), 脳活動から特徴量に変換するデコーダーを訓練した.

テストデータとして, ShapeNet[3] を用いて刺激画像をレンダリングし, その画像を観察中の被験者の脳活動を fMRI により計測し, 全視覚皮質を対象として解析した. また DNN として, ImageNet[4] で 1,000 個のカテ

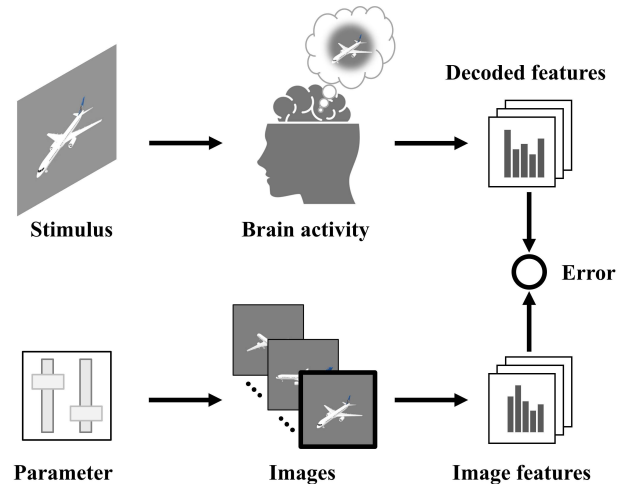


図 1: 実験手法の概要

ゴリに画像を分類するように事前学習した VGG19 モデル [5] を使用した.

図 1 の下段は, 特定のパラメータ値で生成した画像から特徴量を抽出する処理を表す. ここでは, 被験者に提示した刺激画像に対応する 3 次元モデルに対して, 視点を 1 軸周りに 1 度刻みで 360 度回転させてレンダリングし, 特徴量を抽出した. その後, 脳活動からデコードした特徴量との間の誤差 E を以下の式で求めた.

$$E(P) = \sum_{l \in L} \beta_l \|\Phi^{(l)}(\mathbf{v}_P) - \mathbf{y}^{(l)}\|_2^2 \quad (1)$$

ここで, 推定するレンダリングパラメータを P , 対応する CG 画像を \mathbf{v}_P ($224 \times 224 \times 3$) とし, L は DNN の層の集合, $\Phi^{(l)}(\mathbf{v}_P)$ は画像を VGG19 に入力した際の l 番目の層の抽出ベクトル, $\mathbf{y}^{(l)}$ は脳から予測した VGG19 の l 番目の層の特徴ベクトルである. また, β_l は VGG19 の l 番目の層の寄与を決定する重みであり, $1/\|\mathbf{y}^{(l)}\|_2^2$ を VGG19 の各層に対して正規化した値とする. このようにして計算した誤差 E が最小になる回転角度 P を求めた.

また, 脳活動による角度推定に加えて, 刺激画像および位相ランダム画像を用いて同様の角度推定を行い, 各

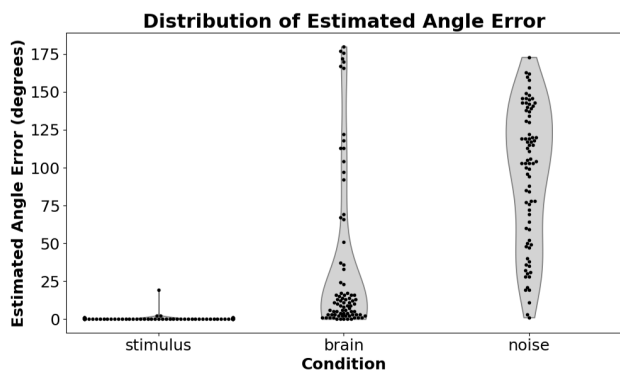


図 2: 推定された回転角度と真値との誤差分布

画像の物体姿勢の推定しやすさを評価した。刺激画像は、異なる姿勢で同じ画像特徴が現れていないかを確認する目的で、位相ランダム画像は、画像の方位・周波数情報のみでどの程度推定可能かを評価する目的で用いた。これらの画像を VGG19 に入力して得られた特徴量を基に、誤差 E が最小になる回転角度 P を推定した。位相ランダム画像の生成については、刺激画像をフーリエ変換し、振幅成分を保持したまま位相成分をホワイトノイズに置き換え、逆フーリエ変換することで生成した。以上、本研究では、ShapeNet20 カテゴリから各 4 枚ずつの画像を用い、計 80 体の 3 次元モデルに対して、刺激画像、脳活動、位相ランダム画像に基づいて回転角度の推定を行い、それぞれの推定精度を比較・評価した。

4 結果と考察

図 2 に、80 体の 3 次元モデルに対して、刺激画像の生成に用いた真の回転角度と、推定された回転角度との間に生じた誤差の分布を示す。ここで回転角度誤差とは、推定角度と真の角度との差の絶対値である。ただし、180 度を超える場合には、360 度からその差を引いた値を誤差と定義する。図 2 より、刺激画像を用いた場合、誤差は 0 度付近に集中しており、推定が極めて正確であることが分かる。一方、位相ランダム画像に基づく推定では、誤差が広範囲に分布しており、推定精度が大きく低下した。脳活動を用いた場合には、一部に大きな誤差が見られるものの、多くのモデルで真値に近い角度が推定されている。位相ランダム画像による推定結果と合わせて考えると、脳活動を用いた場合には、画像の周波数・方位成分以上の情報も推定に利用されている可能性がある。

推定精度の傾向をより詳細に検討するため、図 3 に、刺激画像、脳活動、および位相ランダム画像に基づいて推定した回転角度で生成した画像を示す。図中の (a) 飛行機や (b) 人間の例では、脳活動を用いて刺激画像に近い画像を得ることができた。一方で、(c) のライフルの



図 3: 各推定条件（刺激画像: Stimulus, 脳活動: Brain, 位相ノイズ画像: Noise）における回転角度の推定結果の例。(a) 飛行機モデル, (b) 人間モデル (c) ライフルモデル。

ように、180 度回転させた場合に見た目がほとんど変化しない対称性の高いモデルに対しては、推定された回転角度と真値との間に大きな誤差が見られるものがあった。

5 まとめと今後の課題

本研究では、刺激画像を観察している際の脳活動から、画像生成に用いた回転角度を推定し、その精度を評価した。実験の結果、回転に対して対称性の高いモデルでは推定精度の低下がみられたが、多くのモデルにおいて真値に近い角度を推定できた。今後の課題としては、回転角度だけでなく、照明や材質など他のパラメータについても同様に推定精度の評価をすることが挙げられる。

参考文献

- [1] Horikawa, T., & Kamitani, Y. (2017). Generic decoding of seen and imagined objects using hierarchical visual features. *Nature communications*, 8(1), 15037.
- [2] Shen, Guohua, et al. "Deep image reconstruction from human brain activity." *PLoS computational biology* 15.1 (2019): e1006633.
- [3] Chang, A. X., Funkhouser, T., Guibas, L., Hahner, P., Huang, Q., Li, Z., ... & Yu, F. (2015). ShapeNet: An information-rich 3d model repository. arXiv preprint arXiv:1512.03012.
- [4] Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009.
- [5] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.